Оптимизация параметров алгоритма автоматического распознавания языка речевого сообщения

Леднов Д.А., *кандидат технических наук, с.н.с.,*

Главатских И.А., *ООО «Стел–Компьютерные системы»,*

Ромашкин Ю.Н., *кандидат технических наук*

Рассматривается алгоритм автоматического распознавания языка речевого сообщения на основе параллельной работы нескольких фонетических распознавателей. Проводится оптимизация параметров алгоритма. Излагаются результаты экспериментов по распознаванию русского и ряда иностранных языков.

• распознавание языка • фонетические распознаватели • равновероятная ошибка

The automatic language recognition algorithm based of the PPRLM is considered. Parameters of this algorithm are optimized. The experiment results for Russian and a number of foreign languages recognition are outlined.

• language recognition • phonetic recognizers • equal error rate

Введение

Задача автоматического распознавания языка состоит в определении конкретного языка анализируемого речевого сообщения или принятии решения, что он неизвестен.

Работа алгоритмов распознавания языка основывается, как правило, на сравнении характеристик тестируемого образца речи с ранее обученными моделями различных языков, и результат такого сравнения имеет вероятностный характер. При формировании моделей языков используются фонетико-лингвистические [1] или акустические характеристики речи [2]. Алгоритм, рассматриваемый в данной статье, применяет первый подход. Решение о распознавании некоторого языка принимается конечным классификатором. Он работает на базе метода опорных векторов (Support Vector Machines — SVM) и предварительно обучен на проверяемые языки речи.

Цель статьи — оценка влияния различных параметров алгоритма автоматического распознавания языка на качество его работы.

Описание алгоритма

Реализованный алгоритм определения языка построен по так называемой схеме PPRLM с параллельными фонетическими распознавателями (ФР). Каждый из них, с использованием скрытых марковских моделей и алгоритма Витерби, осуществляет поиск наилучшей последовательности фонетических элементов для анализируемого речевого сообщения на неизвестном языке [3, 4]. Состав ФР в PPRLM схеме постоянен и не зависит от распознаваемого языка. ФР отличаются обучающими выборками, на которых были предварительно обучены, и набором фонетических элементов, которые могут воспроизводить, используя единый международный фонетический алфавит. В качестве вектора информативных признаков для декодирования речевого сообщения используются коэффициенты PLP (Perceptual Linear Prediction) [5], вычисляемые на сегментах фиксированной длительностью 25 мс.

Математическая модель языка является триграмной. Она характеризует вероятность появления той или иной триграммы в данном языке. Для оценки вероятности появления n-грамм, которые отсутствуют в обучающей выборке, используется метод сглаживания вероятностей, разработанный Катцем [6]. При анализе речевого сообщения каждый декодер находит наиболее вероятную последовательность фонетических элементов для используемой модели языка.

Решение о принадлежности распознанной последовательности фонетических элементов тому или иному языку принимается SVM-классификатором. Входным вектором для него служат перплексии, вычисленные для всех используемых ФР:

$$\Pr(\mathbf{x}) = \frac{1}{N}\ln p(x_1, x_2, ..., x_N) = \frac{1}{N}\ln\prod_{i=1}^{N}\ln p(x_i) = \frac{1}{N}\sum_{i=1}^{N}\ln p(x_i),$$

где $p(x_i)$ — вероятность появленния триграммы x_i (в предположении их взаимной независимости), N — общее количество триграмм в исследуемой последовательности фонем.

Сравнение получаемого результата с порогом, установленным пользователем, позволяет принять итоговое решение. Точность принимаемого решения зависит от качества работы каждого из используемых ФР и конечного SVM-классификатора.

Условия и результаты экспериментов

Исследовались русский и 10 иностранных языков: английский (американский), арабский, испанский (европейский), китайский (мандарин), литовский, немецкий, польский французский, турецкий и японский. Для них были предварительно обучены все соответствующие ФР.

Речевые сообщения на русском, литовском и польском языках были собраны самостоятельно в ходе предыдущих работ, а для остальных языков взяты из баз данных CallHome и GlobalPhones. Длительности обучающих выборок при формировании моделей языков представлены в табл. 1.

Таблица 1

Длительности обучающих выборок (Т)

Язык	Т, час
Испанский	48
Английский	35
Арабский	33
Русский	32
Французский	22
Польский	20

Окончание табл. 1

Литовский	18
Турецкий	11
Китайский	7
Японский	6
Немецкий	4

В ходе экспериментов исследовалось влияние на эффективность распознавания следующих факторов:
— объём обучающей выборки;
— размерность GMM-модели фонетических элементов языка;
— количество ФР в PPRLM схеме алгоритма.

Точность работы ФР характеризовалась оценкой вероятности правильного распознавания фонем (FRR)

$$FRR = \frac{N_D}{N},$$

где N_D — количество правильно распознанных фонем в тестовой выборке, N — общее количество фонем в ней. Для каждого языка были сформированы тестовые речевые сообщения длительностью примерно 3 часа, достаточно сбалансированно представляющие фонемный состав каждого языка.

Точность распознавания языка характеризовалась оценкой равновероятной ошибки (EER), при которой ошибки первого и второго родов принимают одинаковые значения (путём подбора порога принятия решения). При этом для каждого языка использовалось не менее чем по 100 фонограмм речи. Их суммарная длительность составила около 66 часов.

Результаты экспериментов представлены в таблицах 2–5.

В табл. 2 на примере ФР испанского языка приведены данные, показывающие влияние объёма обучающей выборки на точность распознавания. Тестовые выборки содержали от V = 100 до 3800 реализаций каждой фонемы языка. Зависимости, полученные для ФР других языков, в целом имели аналогичный характер и поэтому в статье не приводятся. Как видно, с увеличением числа реализаций каждой фонемы в обучающей выборке вероятность правильного распознавания фонемы очень быстро стабилизируется, а равновероятная ошибка распознавания языка сохраняет тенденцию к незначительному уменьшению. Для обеспечения инвариантности ФР к речи различных дикторов и её смысловому содержанию обучающая выборка должна содержать не менее V = 500 реализаций каждой фонемы данного языка. При дальнейшем увеличении обучающей выборки в несколько раз наблюдается несущественный прирост точности работы ФР.

Таблица 2

Влияние объёма обучающей выборки

V	FRR,%	EER,%
100	59,9	14,2
500	61,6	10,5
1000	61,3	10,5
2000	61,1	10,4
3000	60,9	9,8
3800	61,0	10,2

В табл. 3 на примере ФР французского языка показаны экспериментальные данные, отражающие влияние размерности GMM-модели фонемы, которая варьировалась от M = 12 до 100. Они показывают, что с увеличением этой размерности вероятность правильного распознавания фонемы монотонно возрастает. Однако ошибка распознавания языка убывает при этом медленно и, начиная с M = 50, её можно считать практически постоянной. Следует отметить, что с увеличением размерности время обработки также растёт, увеличиваясь примерно в 3 раза при переходе от M = 12 к M = 100.

Таблица 3

Влияние размерности GMM-модели фонемы

M	FRR,%	EER,%
12	67,4	17,1
32	70,3	15,7
50	71,7	15,1
64	72,0	15,1
100	73,0	15,4

Табл. 4 содержит проранжированные данные, характеризующие точность каждого из ФР (при распознавании языка, совпадающим с языком ФР). Прямой связи с длительностями обучающих выборок, указанными в таблице 1, в целом не прослеживается.

Табл. 5 отражает влияние увеличения количества ФР, применяемых в PPRLM схеме алгоритма, когда каждый новый ФР добавлялся в порядке полученного выше ранга.

Таблица 4

Результаты ранжирования ФР

Язык ФР	EER,%
Испанский	10,4
Турецкий	11,1
Арабский	13,3
Польский	13,9
Русский	14,1
Немецкий	14,3
Японский	14,9
Китайский	15,1
Французский	15,7
Английский	15,7
Литовский	15,9

Таблица 5

Влияние количества ФР в алгоритме

Количество ФР	EER,%
1	10,5
2	7,3
3	6,7
4	5,5
5	5,4
6	5,5
7	4,9
8	4,9
11	4,7

Видно, что с увеличением количества используемых ФР равновероятная ошибка распознавания языка практически монотонно уменьшается. При этом, однако, время принятия решения эквивалентно возрастает, увеличиваясь примерно в 4,4 раза при использовании всех 11 ФР.

Заключение

Качество работы алгоритма автоматического распознавания языка зависит от точности каждого фонетического распознавателя, входящего в его состав. Более высокая точность ФР одного языка относительно ФР другого языка не гарантирует более высокую точность автоматического распознавания языка. Актуальным остаётся повышение точности работы всех ФР.

Увеличение количества параллельных ФР в алгоритме приводит к монотонному уменьшению равновероятной ошибки распознавания языка. Однако скорость обработки при этом также падает.

Литература

1. *Аграновский А.В., Зулкарнеев М.Ю., Леднов Д.А., Можаев О.Г*. Автоматическая идентификация языка // Искусственный интеллект, № 4, 2002, изд. НАН Украины, Донецк, 2002. С. 142–150.

2. *Schultz T., Jin Q., Laskowski K., Tribble A. and Waibel A*. Speaker, Accent, and Language Identification Using Multilingual Phone Strings. HLT 2002, San Diego, California, March 2002.

3. *Young S., et al*. The HTK Book (v3.0). Cambridge Univ. Engineering Department. 2000.

4. *Моттль В.В., Мучник И.Б*. Скрытые марковские модели в структурном анализе сигналов. М.: Физматлит, 1999.

5. *Hermansky H*. Perceptual linear predictive (PLP) analysis of speech // Journal of Acoustic Society of America, 1990. Vol. 87(4). P. 1738–1792.

6. *Slava M. Katz*. Estimation of probabilities from sparse data for the language model component of a speech recognizer // IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP35(3):400–401, March 1987.

Сведения об авторах:

Леднов Дмитрий Анатольевич —
кандидат технических наук, старший научный сотрудник, научный консультант научно-технического департамента ООО «Стэл — Компьютерные Системы», основные научные интересы лежат в областях моделей обработки данных, случайных процессов, распознавания речи и идентификации дикторов

Ромашкин Юрий Николаевич —
кандидат технических наук, окончил Московский инженерно-физический институт, факультет «Автоматика и электроника». Область научных интересов: цифровая обработка речевых сигналов, фильтрация речи на фоне помех, автоматическое распознавание речи и языка, идентификация говорящего по голосу, низкоскоростное кодирование речи, оценка качества трактов речевой связи.
E-mail: romayn@yandex.ru

Главатских Игорь Александрович —
ООО «Стэл — Компьютерные Системы», окончил Удмуртский государственный университет. Область научных интересов: речевые технологии — идентификация языка, идентификация диктора, распознавание и синтез речи.
E-mail: ia_glavatskih@stel.ru